AN EXAMINATION OF ALTERNATIVE VARIANCE ESTIMATORS FOR THE MEDICAL EXPENDITURE PANEL SURVEY

Laura Ozcoskun[1] and Samson A. Adeshiyan

United States Bureau of the Census
4600 Silver Hill Road, Washington, DC 20233

Introduction

The Medical Expenditure Panel Survey – Insurance Component (MEPS-IC) is a stratified one-stage sample design that employs an iterative multi-stage weighting procedure that accounts first for unit non-response before post-stratifying to outside control totals. This weighting procedure is both time-consuming and resource intensive, with twenty-one separate iterations. Currently, variance estimates for MEPS-IC are produced using the method of random groups (10 random groups assigned) and are constructed from the **final** adjusted weights. This "shortcut" approach saves considerable computing resources, but yields positively biased variance estimates. Moreover, while theoretically pleasing, the random group variance estimator is known to have high variance in without-replacement samples and yield quite unstable estimates when the number of sampled observations in each random group is small (e.g., domain estimates) or when there is a high rate of unit or item non-response. Consequently, the MEPS-IC methodologists are interested in pursuing a replicate variance approach that more fully captures the variability in the stratification and iterative weighting procedures. Since the MEPS-IC sample design is highly stratified, the traditional stratified jackknife replicate estimator would appear to be the most appropriate replicate method. Unfortunately, the large sample size combined with the complex and repetitive weighting procedures render this method impossible due to the overwhelming computer-resource demands (Adeshiyan et al, 2007).

The choice of an appropriate replicate variance estimator for the MEPS-IC design is complicated: the majority of the non-certainty strata contain several sampled elements, but a small proportion of the non-certainty strata contain a **single** sampled unit. Using the method of random groups sidesteps the need for collapsing strata for variance estimation. These same random group assignments can be used to obtain more stable variance estimators, such as the delete-a-group jackknife proposed in Wolter (1985, Ch.4, p.183). In its simplest form, delete-a-group jackknife replicates are constructed by sequentially dropping each random group from the total estimate and reweighting the remaining cases by the unconditional probability of remaining in the replicate; this approach is used by several ongoing programs in our Economic Directorate, including the Annual Capital Expenditures Survey and the Quarterly Financial Report. This simple method yields unconditionally unbiased estimates of variances when the estimator itself is **unbiased** and **linear** and when the random groups are selected in the same manner as the parent sample.

Both the delete-a-group jackknife and the method of random groups are unbiased for stratified simple random samples when at least one sampled unit in each stratum is represented in each random group. Kott (2001) presents a variation on the traditional delete-a-group jackknife replicate weighting that allows computation of a bound on the bias of the variance estimator when this condition is not met, provided that the number of sampled units in each stratum is greater than five. Additionally, he presents an "extended delete-a-group jackknife variance estimator" that produces "nearly unbiased" variance estimates under

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

these circumstances. The delete-a-group jackknife replicates proposed in the Kott paper employ strata-specific replicate factors.

The theory that supports both versions of the delete-a-group jackknife is developed for simple expansion (unbiased) estimators of **totals**. Extending this theory to include calibrated estimates of totals is less straightforward, and further extending this theory to ratios of calibrated estimates is not addressed in the literature (as far as we could ascertain). The key statistics released by MEPS-IC fall into this latter category. To make sure that our empirical results dovetail with the published theory for the considered variance estimators, we restrict our analysis to an important MEPS-IC total, assuming that the ratio-estimate variances could easily be obtained via Taylor linearization as described in the Variance Estimation Methodology Section below. Before fully recommending a change in the current procedure, additional research that includes ratio estimation should be done.

The MEPS-IC weighting procedures described below are iterative proportional fitting (raking) techniques as described in Kalton and Flores-Cervantes (2003). The final estimates are not linear, and consequently, the replicate variance estimates are not necessarily unbiased over repeated samples. We attempt to use unbiased replication procedures, following the recommended procedure for a calibrated estimator with the delete-a-group jackknife proposed in Crouse and Kott (2004).

Having chosen a variance estimation method, a second concern must be addressed: Is there a detrimental effect on the variance estimates of <u>not</u> fully replicating an iterative weighting procedure (e.g., replicating three iterations of a procedure instead of seven)? Canty and Davison (1999) explored a similar question with a stratified jackknife estimator and showed the reverse empirically (i.e., the variance estimates constructed from a fully replicated procedure were "less stable" or "more biased" than those constructed from partially replicated procedure.) The question of whether similar findings would hold with the MEPS-IC data was certainly worth pursuing, if only from a time and computer-resource saving perspective.

In this paper, we address two concerns for the MEPS-IC: the appropriate choice of variance estimator and the decision to fully or partially replicate a weighting procedure for the selected procedure. Our research focuses on calibrated expansion estimates. The next section provides background on the MEPS-IC survey design and estimation procedures. Then, we provide an overview of the considered variance estimation methods. Next, we present a simulation study designed to assess the statistical properties of these variance estimators under a slightly simplified design. After that, we present empirical results for all considered methods using 2003 MEPS-IC data. We conclude with recommendations for future MEPS-IC applications and research, along with some general remarks.

Background

The Medical Expenditure Panel Survey - Insurance Component (MEPS-IC) is an annual survey of business establishments (locations) and governments. The survey is funded by the Agency for Healthcare Research and Quality and conducted by the U.S. Census Bureau. Data are collected on various aspects of employer-sponsored health insurance, such as whether health insurance is offered, the number of employees enrolled in health plans, and the premium amounts, including the employee and employer contribution to the premium. MEPS-IC measures quantities such as average premium and contribution per enrollee, the percentage of employees enrolled, and the percentage of establishments and governments contributing to health insurance. Because employers are a key source of health insurance in the United States, data are used by federal agencies, such as the Bureau of Economic Analysis, the Centers for Medicare and Medicaid Services, and the Department of the Treasury, and by state governments, to monitor and predict national and state trends in employer-sponsored health insurance.

MEPS-IC uses two list frames from which two samples are selected: a private sector sample of establishments selected from the Census Bureau's Business Register, and a public sector sample selected from the Census of Governments. By far the largest portion of the sample is from the private sector where the sample size is approximately 42,000 establishments annually. Sommers and Reisz (2003) provide

details on the MEPS-IC stratification procedure. Complete details on the MEPS-IC design and methodology are available at http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr6/mr6.pdf and http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr18/mr18.pdf.

The MEPS-IC final weights are the product of the sampling weight, a unit non-response adjustment factor, and a poststratification factor. The unit non-response adjustment factor is computed iteratively by dividing the sample into disjoint weighting cells defined by eight combinations of establishment size and firm size within the 50 states plus the District of Columbia cross-classified by type of firm (single or multi-establishment), and industry. Cross-classifying all these variables results in a large number of cells with small counts, many of which may be zero. Hence, sampling weights are raked to **marginal** cell totals.

The non-response adjustment for MEPS-IC is performed in two stages. During data collection, sampled cases are contacted by telephone to determine whether they offer health insurance (pre-screener stage). Those that do **not** are considered respondents but are not mailed a form. Those that do offer health insurance are not considered respondents unless they provide further information on the insurance either by mail or telephone. Thus, the first stage in unit non-response weighting compensates for unit non-response to the initial telephone (pre-screener) interview that indicated that they do in fact offer health insurance. The second stage of the unit non-response adjustment accounts for unit non-response where it is unknown if the non-respondent offers health insurance incorporating the weights from the first stage of the adjustment. This two-stage unit non-response adjustment process is fully replicated when we compute the empirical variance estimates. For the simulation variance estimates, we used a one-stage unit non-response adjustment similar to most surveys that have only one stage of data collection.

MEPS-IC uses seven iterations of raking to ensure that the weights converge and that their sums add up to the required marginal cell totals. In our application, we found out that three iterations were sufficient for both full sample and replicate weights after verifying that the adjustment factors had converged in all cells.

The unit non-response adjusted weights are next controlled to independently obtained employment control totals via an iterative two-way raking procedure. This poststratification procedure uses two cell categories: eight combinations of establishment size and firm size within the 51 state sets and industry type. The raking is accomplished in seven and one-half iterations (ending with the state by size group rake). With both the nonresponse and poststratification weighting adjustment procedures, we noticed that both of these adjustment factors converged after one or two of the 7 or 7.5 iterations. So, for this study we consider 4 replicate reweighing procedures for each considered variance method:

- Replicating the above described weighting procedure
- Performing 3 iterations (instead of 7 and 7.5 iterations) for both the nonresponse and poststratification weighting adjustment procedures
- Performing 1 iteration for both the nonresponse and poststratification weighting adjustment procedures (this is only considered in the simulation part of the study)
- Performing no iterations; using final weights to construct the replicate weights.

As stated in the introduction, MEPS-IC currently uses the method of random groups to produce variance estimates. The production variance estimates are constructed from 10 random groups, and the weighting procedure is not replicated. Sampled units are assigned to random groups as part of the sample selection.

In the following sections, we briefly consider the MEPS-IC "shortcut" procedure, but we focus on the possibility of obtaining reasonable variance estimates with a replicated procedure that uses a reduced number of iterations at each weighting stage.

Variance Estimation Methodology

Our discussion below concentrates on expansion estimates of the form $\hat{X} = \sum_h \sum_{j \in h} \widetilde{w}_{hj} y_{hj}$, where $\widetilde{w}_{hj}$ is the final weight associated with unit *j* in stratum *h*. In our applications, the expansion estimates are non-linear, since this final weight is the product of several iterative adjustments. A Taylor linearization estimate of the variance of a ratio estimate ($\hat{R} = \frac{\hat{X}}{\hat{Y}}$) whose numerator is a **proper subset** of the denominator is given by

$$\hat{v}(\hat{R}) \approx \left(\frac{\hat{X}}{\hat{Y}}\right)^2 \left[\frac{\hat{v}(\hat{X})}{\hat{X}^2} - \frac{\hat{v}(\hat{Y})}{\hat{Y}^2}\right]$$

The discussion below focuses on obtaining the necessary input variances for this linearization formula. Note that this linearization formula can easily include fpc-corrections in the variance estimates and does not require an unbiased estimation procedure.

To use the method of random groups, we randomly divide the non-certainty component of the sample into *G* mutually exclusive groups using the same sampling methodology used to select the parent sample (Wolter, 1985, pp. 31-32). Each random group's sample weight (w_{hj}) is then reweighted to represent the full sample by multiplying the random group estimate by *G*. [Note: *G* sets of replicate weights are assigned to each sample unit *j*, where the gth replicate weight is zero when unit *j* is in random group *g*]. Certainty cases are included in each random group without any replicate weight adjustment. The full sample estimation procedure is then applied to each of the replicate weights (e.g., non-response adjustments, post-stratification). These replicate weights are then used to calculate replicate estimates for the characteristic of interest.

The random group variance for an estimate $\hat{X}$ is

$$v_{RG}(\hat{X}) = \frac{\sum_{i=1}^{G} (\hat{X}_i - \hat{X}_0)^2}{G(G-1)}$$

where *G* is the number of random groups, $\hat{X}_i$ is the ith replicate estimate, and $\hat{X}_0$ is the full-sample estimate.

Random group variance estimation has two drawbacks. First, random group estimation can be unpredictable when applied to samples selected without replacement because the random group estimator "tends to estimate the variance as if the sample were selected with replacement" (Wolter, 1985, p.43). The second drawback is the instability of the random group variance estimates, especially when the number of sampled observations in each random group is small (as with the MEPS-IC sample) or when there is a high percentage of unit non-response.

The delete-a-group jackknife variance estimation method can be applied to the same types of sample designs as the random group method. Again, the non-certainty portion of parent sample is divided into *G* random groups. However, the delete-a-group jackknife replicate estimate is computed for each replicate *g* by removing the gth random group from the full sample. As with random group estimates, jackknife replicates are obtained by weighting each replicate estimate to represent the full sample. Again, certainty cases are included in each replicate with no replicate weight adjustment. Since jackknife replicate sample sizes are larger than the corresponding random group replicate sample sizes, delete-a-group jackknife variance estimates are often more stable at least for smooth statistics such as expansion estimators, ratio estimators, or regression estimators. The delete-a-group variance estimator for an estimate $\hat{X}$ is

$$v_{DAG}(\hat{X}) = \frac{(G-1)}{G}\sum_{i=1}^{G}(\hat{X}_i - \hat{X}_0)^2 \qquad (1)$$

where G is the number of delete-a-group jackknife groups, $\hat{X}_i$ is the ith delete-a-group jackknife group estimate, and $\hat{X}_0$ is the full-sample estimate.

As stated in the introduction, MEPS-IC employs a calibration estimator, with iterative proportional fitting used at both the unit-nonresponse and poststratification weighting stages. For our replication, we follow the recommended procedure outlined in Crouse and Kott (2004) for both random group and delete-a-group jackknife estimation. We refer to this variance estimator as the simple delete-a-group jackknife estimator.

Kott (2001) presents a conditionally unbiased[2] delete-a-group jackknife variance estimator for a stratified SRS-WOR design that employs strata-specific replicate factors defined as

$$n_h / (n_h - n_{hg}) \qquad (2)$$

where n_h denotes the number of sampled units in stratum h and n_{hg} the number of sampled units assigned to random group g in stratum h. For a stratified SRS-WOR design and a **simple expansion estimator**, Kott (2001) proves that his delete-a-group jackknife variance estimator is approximately unbiased for the true variance when (1) the sample size in each stratum is larger than the number of random groups and (2) all sampling fractions are negligible (less than or equal to 1/5) and is biased upwards otherwise. Thus, the delete-a-group jackknife estimator is unbiased if and only if units from each sample stratum are represented in each replicate. For a simple unbiased total estimate, Kott (2001) develops the **extended** delete-a-group jackknife to account for the situation where condition (2) is true and condition (1) is not, specifically where $n_{hg} = 0$ in several strata. Let S_{hg} be the set of n_{hg} sample units in stratum h and random group g. The extended delete-a-group jackknife (DAGE) replicate weights are

$$w_h^g \begin{cases} w_{hj} \text{ when } S_{hg} \text{ is empty} \\ w_{hj}(1-[n_h-1]Z) \text{ when } j \in S_{hg} \\ w_{hj}(1+Z) \text{ otherwise} \end{cases} \qquad (3)$$

where $Z^2 = G/[(G-1)n_h(n_h-1)]$. The extended delete-a-group jackknife reduces to the "usual" delete-a-group jackknife variance estimator when all n_h are greater than G.

Note that the extended delete-a-group jackknife requires at least **two sampled** elements per stratum. Unfortunately, in more than one instance, the MEPS-IC sample comprises one sampled unit from a stratum. Consequently, for extended delete-a-group jackknife replication, such strata must be collapsed; note that this same problem occurs with a stratified jackknife application. Our extended delete-a-group jackknife begins with the replicate weights assigned in equation (3), then calibrates the replicate sampling weights using the Crouse and Kott (2004) procedure. Kott (2001) does not provide any theoretically-based modifications for the **extended** delete-a-group jackknife variance estimator with a calibrated estimator, and the Crouse and Kott (2004) paper does not employ the extended delete-a-group jackknife. Thus, our application follows the procedures described for the simple delete-a-group jackknife estimator after assignment of replicate weights (using the extended delete-a-group jackknife replicate weights), and this "ad hoc" (and not theoretically driven) approach is definitely a concern in interpreting our results.

In theory, the extended delete-a-group estimator should be better suited to the MEPS-IC sampling design than the other variance estimators considered in this paper. However, the DAGE replicate factors are defined under the assumption of stratified SRS-WOR sampling design, where the design weights are equal to the inverse of the probability of inclusion. This requirement is <u>not</u> strictly observed in the MEPS-IC

[2] conditioned on the selected sample.

production system, although it is in our simulation study. Moreover, the extended delete-a-group jackknife requires that each stratum contain at least two sampled elements per stratum, but this condition is not met with the MEPS-IC design. So, this requires that the survey practitioner employ a collapsed stratum procedure, which can add bias to the variance estimates.

Simulation Study

Our evaluation begins with a simulation study designed to assess variance estimation affects using the alternative replicate estimators. Our study uses a simplified MEPS-IC design drawn from a simulated population. For our simulation, we confined our population to four industries proposed by our subject-matter experts (Construction, Retail Trade, Finance, and Wholesale Trade) in four states (Illinois, New York, North Carolina, and Washington). Instead of using the survey's sampling strata, we created three collapsed firm-size categories in each region *x* industry cell.

We modeled our population data from the 2003 MEPS-IC empirical data, using the SIMDAT algorithm (Thompson, 2000) to generate multivariate observations of current and previous year's establishment employment size within each possible industry/state/size class combination, matching the original sample population counts.

From this population, we selected 5000 stratified simple random samples without replacement using the same sampling rates as the original survey. In each sample, we randomly assigned response status indicator variables and multi-establishment/single-establishment indicator variables to each sampled unit using propensities modeled from our survey data.

We used these 5000 random samples to construct the empirical variance ("truth") of our final weighted estimate of total employment:

$$V_T(\hat{X}) = \frac{\sum_{r=1}^{5,000} (\hat{X}_r - \overline{X})^2}{5,000}$$

where $\hat{X}_r$ is the rth sample estimate, and $\overline{X}$ is the mean of the 5000 sample estimates. Our fully calibrated estimator from these 5,000 samples is essentially unbiased, i.e., $(\overline{X} - \mu)/\mu = 0.01,$ where μ is the population value of total employment.

In 1,000 of the 5000 samples, we assigned sample units to **16** random groups, then used these 1000 samples to study the statistical properties of each variance estimation method for this estimator over repeated samples, computing the following statistics:

Relative Bias

$$\frac{\frac{1}{1,000}\sum_{r=1}^{1,000} v_{meth}(\hat{X}_r)}{V_T(\hat{X})} - 1$$

where $v_{meth}(\hat{X}_r)$ is the variance estimate for a given variance estimation method and sample *r,* and $V_T(\hat{X})$ is the "true" variance.

Stability

$$\frac{\sqrt{\frac{1}{1,000}\sum_{r=1}^{1,000}[v_{meth}(\hat{X}_r) - V_T(\hat{X})]^2}}{V_T(\hat{X})}$$

Coverage The proportion of 90% confidence intervals constructed from each of the 1,000 samples/variance estimator (*meth)* that contain the true population total. Following Kott (2001), our confidence intervals are constructed with a *t*-statistic with 15 degrees of freedom.

Relative bias measures the proximity of the variance estimator to the true variance over repeated samples, as well as the direction of the bias. The stability measures the variability of the variance and can be viewed as a coefficient of variation of the variance estimators. The optimal variance estimator will have relative bias and stability near zero, and 90% confidence interval coverage. The number of random groups is a key difference between our simulation study and the realized design. For our simulation study, we decided against using ten random groups (as done with the MEPS-IC survey) because of the decrease in degrees of freedom in computing coverage rates, believing that the coverage rates constructed with a *t*-statistic with nine degrees of freedom could potentially be too high to distinguish differences between any of the variance estimators.

Table 1 presents our simulation results. All statistics are reported as percentages. We present four sets of results per variance estimation methods: one that uses the final weights to construct replicate estimates (0 iterations); two that utilizes "partially" replicated weighting procedures (1 and 3); and one that fully replicates the production weighting procedures (7). Coverage rates that are **not** statistically different from the nominal 90-percent are indicated by an asterisk.

Table 1: Simulation Results

Variance Estimation Method	Estimate	Number of Iterations	Relative Bias	Stability	Coverage
Random Groups	Total Employment	0	184.03	214.42	100
		1	9.13	643.87	94.5
		3	16.89	577.91	95.4
		7	21.96	566.38	96
Delete-a-Group Jackknife	Total Employment	0	184.03	214.42	100
		1	-34.88	45.89	94.3
		3	-37.70	47.76	93.8
		7	-37.67	47.75	93.8
Extended Delete a Group Jackknife	Total Employment	0	1444.33	1470.08	100
		1	-44.73	51.86	92.6
		3	-45.76	52.59	91.3*
		7	-45.71	52.55	91.3*

The results from Table 1 can be summarized as follows:

- Regardless of variance estimator, the variance estimates constructed with a shortcut estimator are highly positively biased and extremely unstable. These effects are particularly evidenced with the extended delete-a-group jackknife variance estimator. This provides some evidence for discontinuing the current production procedure, which constructs replicate weights from final weights.
- With the method of random groups, the positive bias (overestimation) of the variance estimates increases as iterations are added to the replication procedure. This has the undesirable effect of increasing the overestimation as the replication procedure more closely mimics the production procedure. Moreover, these estimates are very unstable. The confidence intervals constructed with the

random groups estimates are very wide. These wide intervals would make it difficult to test for significant differences between parameters of interest. Again, these results provide evidence for discontinuing the current production procedure variance estimation method for MEPS-IC.

- In contrast to the random group variance estimates the simple delete-a-group jackknife estimates are underestimates. However, they are considerably more stable than their random group counterparts. With this set of samples the degree of underestimation does not appear to greatly affect coverage; all confidence intervals are still conservative, although improved over the random group intervals, and the coverage rates for three and seven iterations are closer to the nominal value of 90%. Both the relative bias and stability worsen slightly as the number of iterations increases.
- The extended delete-a-group jackknife variance estimates appear to be less sensitive to the number of iterations used than the other methods. Like their simple delete-a-group jackknife counterparts, these variance estimates are quite negatively biased. Moreover, their stability is slightly higher, reflecting the additional variability due to the replicate factors. Coverage rates for three and seven iterations are at the nominal value, demonstrating very little practical impact on coverage from the underestimation.

Ultimately, the simulation study results demonstrate that some form of weight adjustment replication is clearly preferable to the currently used shortcut procedure, regardless of variance estimator. With all methods, the results from a partially replicated weighting procedure (using three iterations) have similar statistical properties as those obtained from a fully replicated procedure. This result is quite consistent with those presented in Canty and Davison (1997).

Considered jointly the simulation study results provide evidence that both versions of the delete-a-group jackknife estimator have important statistical advantages over the method of random groups. These advantages are particularly evident in terms of stability. Both versions of the delete-a-group jackknife estimator yield estimates that are drastically more stable than their random groups counterparts. The advantage of using the more stable variance estimator is evidenced by the improved coverage rates. Certainly the degree of underestimation shown by the relative bias is an area of concern. We suspect that these statistics are exaggerated by the results of few extreme samples. (This exaggeration applies to the random group bias estimates as well).

The choice of simple versus extended delete-a-group jackknife is less obvious when viewing the simulation study results. With three iterations, the stability is quite comparable. The simple delete-a-group jackknife is less biased than its extended delete-a-group jackknife counterpart, but both are severe underestimates (with an essentially unbiased estimator). The coverage rates for the extended delete-a-group tend to "tip the scales" towards this method, but with some cautions in interpretation, recognizing that we do not fully replicate the survey design and cannot be completely sure the differences between our model and the true data are sufficiently substantial to effect our results.

Empirical Results

Survey decisions are rarely made based on simulation results alone. Table 2 presents empirical comparisons constructed from the 2003 MEPS-IC data set. We include results from the current production method (random groups/0 iterations) for comparative purposes only.

Table 2: Empirical Results (2003 MEPS-IC Data)

Variance Estimator	Number of Iterations	Estimate	Estimate	Standard Error	Coefficient of Variation
Random Groups	0	Total Employment	109576005	1,651,697	0.0151
	3	Total Employment	109576005	643,427	0.00587

	7	Total Employment	109576005	382,860	0.0035
Delete-a-Group Jackknife	3	Total Employment	109576005	296,079	0.0027
	7	Total Employment	109576005	311,424	0.0028
Extended Delete-a-Group Jackknife	3	Total Employment	109576005	546,475	0.00499
	7	Total Employment	109576005	483,274	0.00441

In interpreting these empirical results, it is important to remember that they are computed from one sample, not one thousand and that we use ten random groups instead of 16. As a result, the extended delete-a-group jackknife replicate factors are generally the stratum-specific replicate factors described by (2). The shortcut standard error is relatively large which implies a wide confidence interval. This observation is similar to those from the simulation study results. Finally, notice that the simple delete-a-group jackknife standard errors appear to be less sensitive to the number of weighting iterations than the other methods. This is consistent with the simulation study results.

Conclusion

Right now, MEPS-IC uses the method of random groups with replicate weights constructed from the final weights to estimate their variances. The many statistical drawbacks of this approach were known beforehand, but it was believed that the degree of standard error overestimation was not severe. Our report presents evidence to the contrary for total estimates. Also, results from our simulation study provide evidence against the usage of the random group variance estimator since the variance estimators are unstable even when the weight adjustment procedure is replicated.

Where does this take us? Again, from our simulation study using a totals estimates it appears that there are substantial statistical benefits to replacing the random group estimator with one of the two examined delete-a-group jackknife variance estimators. The choice of which version is less obvious.

The results from the simulation study tend to support the use of the extended delete-a-group jackknife over the simple jackknife in terms of confidence interval coverage. On the other hand, the variance estimates constructed from the simple delete-a-group jackknife have smaller absolute bias and are slightly more stable then their extended delete-a-group jackknife counterparts. Both sets of corresponding variance estimates are underestimates, and the degree of underestimation on the average from either method is not insubstantial. However, both sets of delete-a-group jackknife variance estimates are fairly stable, so it is quite possible that the large relative biases are caused by a few very divergent samples (this contention is further supported by confidence interval coverage rates that are close to the 90% nominal value obtained with either variation of the delete-a-group jackknife).

The empirical results are inconclusive. We believe that it would be unwise to make any decisions about the MEPS-IC variance estimation procedure based on the empirical results alone or to extrapolate these results to other samples. On paper, the extended delete-a-group jackknife variance estimator is more suited to the MEPS-IC design. Unfortunately, the realized MEPS-IC sample does not satisfy all of the assumptions required for unbiased variance estimation with this method. Furthermore, we use an ad hoc application of the extended delete-a-group jackknife, whose theory is only fully developed for a simple expansion estimator, not a calibrated estimator. In contrast, there is some theory to support the simple delete-a-group jackknife variance estimator with a calibrated estimator, at least for survey totals.

As just stated, our research was limited to survey total estimates, but the key estimates for MEPS-IC are ratio estimates. Before making any final recommendations for MEPS-IC similar research needs to be repeated employing these key ratio estimates. With that said, when considering MEPS-IC total estimates, the only clear choice from our presented study is against any form of the random group variance estimator and against the usage of a "shortcut" variance estimator that uses final weights to construct the replicate weights. Our results suggest that either method of the delete-a-group jackknife variance estimator would be preferable, and that we can achieve comparable results without **fully** replicating the weighting procedure. We cannot say that either delete-a-group jackknife variance estimator is anywhere near optimal for this survey design and calibration procedure.

Finally, we continue our quest for a variance estimator for MEPS-IC that has good statistical properties, is consistent with the survey design, and is not overly resource-intensive. Adeshiyan et al (2007) eliminated the stratified jackknife due to resource constraints but were unable to find a consistently good linearized jackknife estimator, so these methods are not considered (they probably would be anyway, given the notable proportion of non-certainty strata containing one sampled unit). With this survey's design, one could consider using the bootstrap as an alternative replicate variance estimator. An alternative approach could be to use a model-assisted variance estimator as proposed in Deville and Sardnal (1992) or the design-based (non-replicate) variance estimator described in Lu and Gelman (2003). The latter approach is particularly appealing, since the authors' apply their variance estimation decomposition to a survey with a very similar design and a similarly complex iterative multi-step weighting procedure to MEPS-IC. In the meantime, we recommend the standard or extended delete-a-group jackknife variance estimation over the method of random groups.

Acknowledgements

The authors thank Anne Kearney, Rita Petroni, and Jenny Thompson for their useful comments on earlier versions of this manuscript.

References

Adeshiyan, S., Thompson, K. and Ozcoskun, L. (2007, to appear) Investigation of Jackknife Linearization Variance Estimators for US Census Bureau Business Survey Estimates. *Proceedings of the Third International Conference on Business Surveys*, American Statistical Association.

Canty, A.J. and Davison, A.C. (1999). Resampling-based Variance Estimation for Labour Force Surveys. *The Statistician*, **48**, pp. 379-391.

Crouse, C.and Kott, P.S. (2004). Evaluating Alternative Calibration Schemes for an Economic Survey with Large Nonresponse. *Proceedings of the Survey Research Methods Section,* American Statistical Association.

Deville, J.C. and Sarndal, C.E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, **87**, pp. 376-382.

Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *Journal of Official Statistics*, **19**, pp. 81-97.

Kott, P. (2001). The Delete-A-Group Jackknife. *Journal of Official Statistics,* **17**, pp. 521-526.

Lu, H. and Gelman, A. (2003). A Method for Estimating Design-Based Sampling Variances for Surveys with Weighting, Poststratification, and Raking. *Journal of Official Statistics*, **19**, pp. 133-151.

Sommers, J. and Riesz, S. (2003). A New Stratification for the Medical Expenditure Panel Survey – Insurance Component (MEPS-IC). *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 3987 – 3992.

Thompson, James R. (2000). Simulation: A Modeler's Approach. New York: John Wiley & Sons, pp. 87—110.

Wolter, Kirk M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag, Inc.